ISRN TELCORDIA--2012-06+PR-0GARAU
Prepared for: Office of Naval Research

Fast Multiscale Algorithms for Information Representation and Fusion

Technical Progress Report No. 6

Devasis Bassu, Principal Investigator

Contract: N00014-10-C-0176

Applied Communication Sciences

One Telcordia Drive

Piscataway, NJ 08854-4157

January 2012

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JAN 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE **Fast Multiscale Algorithms For Information Representation And Fusion**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Applied Communication Sciences,One Telcordia Drive,Piscataway,NJ,08854**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**12**	

1 Abstract

In the sixth quarter of the work effort, we focused on a) conducting experiments on real-world data sets using the developed algorithms, b) design/implementation of the Multiscale Singular Value Decomposition (SVD) algorithm and c) fine tuning and bug fixes for the randomized SVD and ANN algorithms. This report documents the current variants of the Multiscale SVD algorithms under development.

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and continue improving the multiscale SVD algorithms. No problems are currently anticipated.

Table of Contents

2 Summary

In this quarter, we continued design of the new multiscale SVD algorithms. Developement of the algorithms is underway.

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and continue improving the multiscale SVD algorithms. No problems are currently anticipated.

3 Introduction

The primary project effort over the last quarter focused on completing the design of the multiscale SVD algorithms [1]. Descriptions of the multiscale SVD algorithms are provided in Section 4.

4 Methods, Assumptions and Procedures

4.1 Multiscale Singular Value Decomposition

The Singular Value Decomposition (SVD) [2] is a fundamental tool in linear algebra which provides a factorization of any real or complex matrix. It provides complete spectral information for any linear operator. Given an *m* x *n* matrix *A* of rank *k* < min(*m*,*n*), the SVD represents *A* in the form

$$A = U \circ D \circ V^*$$

where *D* is a *k* x *k* diagonal matrix whose elements are non-negative, and *U* and *V* are matrices (of sizes *m* x *k* and *n* x *k*, respectively) whose columns are orthonormal. The compression provided by the SVD is optimal in terms of accuracy [3], and has a simple geometric interpretation: it expresses each of the columns of *A* as a linear combination of the *k* (orthonormal) columns of *U*; it also represents the rows of *A* as linear combinations of (orthonormal) rows of *V*; and the matrices *U*, *V* are chosen in such a manner that the rows of *U* are images (up to a scaling) under *A* of the columns of *V*.

However, for any given data set of observed points, the SVD may not necessarily be locally optimal. This problem discussed in detail in the earlier technical report ISRN TELCORDIA--2011-04+PR-0GARAU. The Multiscale SVD (MSVD) addresses this by providing a spectral readout at all scales. The algorithms for both small and high dimensions are described below.

4.1.1 Small Number of Dimensions (less than or equal to 3)

Here, we consider multivariate data streams that are tagged using a small number of dimensions. An example would be multivariate time-series data comprising readings from several sensors. Note that while the number of sensors may be very large, after registration, each "observed" vector, of possibly high dimensionality, is associated with a scalar variable – in this case, time. If additional spatial information were available, then each vector would be tagged with a 3-dimensional variable, namely (time, latitude, longitude).

Let $\{x_i | x_i \in \Re^n\}$ for $i = 1,2,\dots,N$ represent the data set where each x_i is associated with a vector $y_i \in \Re^d$. We outline the algorithm for small d below; the next section outlines the algorithm for larger values of d. The basic construction of MSVD involves imposing a dyadic grid on a window (hyper-cube) of appropriate size (based on application needs). For static datasets, this may be the entire dataset while for streaming datasets, it could be a sliding or non-overlapping window of a specific size. Without loss of generalization, assume that y_i resides in the unit hypercube $[0,1]^d$.

1. Impose a dyadic grid on the unit hypercube $[0,1]^d$ up to scale S. Define the interval $I_{s,j_1,j_2,\ldots,j_d} \subset [0,1]^d$ where the i-th dimension of $I_{s,j_1,j_2,\ldots,j_d}$ is in $[\frac{j_i}{2^s}, \frac{j_i+1}{2^s})$ for $j_i = 0,1, \ldots, (2^s - 1)$ and $s = 1,2, \ldots, S$.
2. Define $X_{s,j_1,j_2,\ldots,j_d} = \{x_i | y_i \in I_{s,j_1,j_2,\ldots,j_d}\}$ as the subset of data points in the interval $I_{s,j_1,j_2,\ldots,j_d}$.
3. Construct the matrix $M_{s,j_1,j_2,\ldots,j_d}$ using the data points in $X_{s,j_1,j_2,\ldots,j_d}$ with column size n. The number of rows is equals to the number of points in $X_{s,j_1,j_2,\ldots,j_d}$.
4. Compute the SVD for $M_{s,j_1,j_2,\ldots,j_d} = U \circ \Sigma \circ V^*$. Store the singular values and singular vectors $\{(\sigma_1, \sigma_2, \ldots, \sigma_k), v_1, v_2, \ldots, v_k\}$ where v_* is a n-dimensional vector.
5. Repeat steps 2 through 4 for all scales and intervals.
6. For any given interval at the finest scale, there are exactly $(S + 1)$ intervals (one at each scale) that contain it. The corresponding sets of singular values and vectors completely characterize the data cloud for that interval.

The dyadic tree of singular values and vectors constitute the MSVD of the dataset.

4.1.2 High Number of Dimensions (greater than 3)

The primary issue with higher dimensions is the exponential growth in the number of intervals to be processed. Further, the actual measurement data is kept separate from the construction of the dyadic grid in the scheme described above. In the general setup, consider any multivariate dataset normalized approximately to the unit ball. The objective is to create multiscale characterizations by considering balls of sizes $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$, ... around each point in the dataset. The rest of the construction is similar as described earlier. Briefly, construct the SVD for the points in each ball to obtain the MSVD.

For very large datasets, we will use the randomized approximate nearest neighbors algorithm (defined in the earlier technical report ISRN TELCORDIA--2011-03+PR-0GARAU) to obtain random samples of points contained in balls at multiple scales. This provides a rapid way to construct the MSVD tree with the desired scaling behavior.

4.2 Deliverables / Milestones

Date	Deliverables / Milestones	Status
Oct 2010	Progress report for period 1, 1st quarter	✓
Jan 2011	Progress report for period 1, 2nd quarter / complete randomized matrix decompositions task	✓
Apr 2011	Progress report for period 1, 3rd quarter / complete approximate nearest neighbors task	✓
Jul 2011	Progress report for period 1, 4th quarter / complete experiments – part 1	✓
Oct 2011	Progress report for period 2, 1st quarter	✓
Jan 2012	Progress report for period 2, 2nd quarter / complete multiscale SVD task	✓
Apr 2012	Progress report for period 2, 3rd quarter	
Jul 2012	Progress report for period 2, 4th quarter / complete experiments – part 2	
Oct 2012	Progress report for period 3, 1st quarter	
Jan 2013	Progress report for period 3, 2nd quarter / complete multiscale Heat Kernel task	
Apr 2013	Progress report for period 3, 3rd quarter	
Jul 2013	Final project report + software + documentation on CDROM / complete experiments – part 3	

5 Results and Discussion

There are no benchmarks or experimental results to report for this quarter.

6 Conclusions

The project is on track with completed design of the multiscale SVD algorithms along with an initial implementation. We will continue with algorithmic improvements and experimentation using the developed algorithms in the next quarter.

No problems are currently anticipated.

7 References

[1] G. Lerman, *Quanitfying curvelike structures of measures by using Jones quantities*, C.P.A.M., vol. 56, issue 8, pages 1294-1365.

[2] G. H. Golub, W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, vol. **2** (2), pages 205–224, 1965.

[3] G.H. Golub, C.F. Van Loan, *Matrix Computations (3rd edition.)*, Johns Hopkins University Press, Baltimore, 1996.